IB Review # 3

Differential Calculus: Tangent and Normal Lines

Normal Distribution

Chi Square Statistic

Introduction to Differential Calculus

Vocabulary:

Gradient → Slope → $\frac{\Delta y}{\Delta x}$

Derivative → Rate of change anywhere on the curve, which = slope of the tangent line at any point on the curve.

Differentiation

The process of finding the derivative.

Differentiation

for: $y = ax^n$

Derivative: $y' = a \cdot n x^{n-1}$

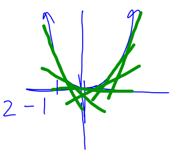
slope of tangent

Example: $y = 2x$

$y' = 2x^0$

$y' = 2$

$y = x^2$



$y' = (1)(2)x^{2-1}$

$y' = 2x$ @ $x = -1$ slope $= -2$

@ $x = 0$ slope $= 2(0)$ $= 0$

@ $x = 2$ slope $= 2(2)$ $= 4$

Tangent and Normal Lines

Find the line tangent to $y = 5x^2 - 2x + 6$

at the point (1, 9).

$y' = 10x - 2$

$y - y_1 = m(x - x_1)$

$y'(1) = 10(1) - 2$
$= 8$

$y - 9 = 8(x - 1)$

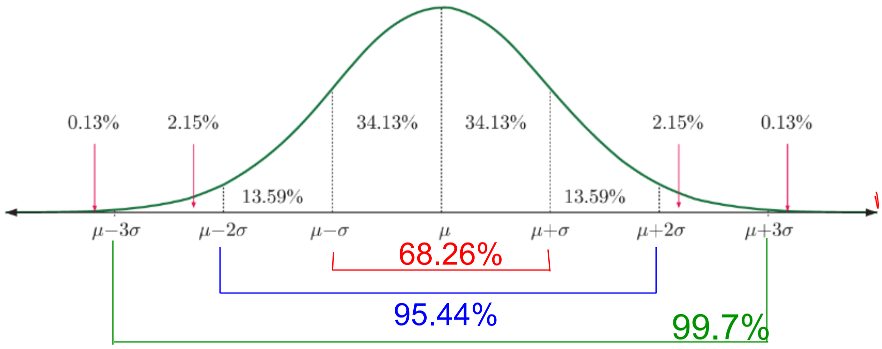
Find the normal line.

Perpendicular to tangent.

$\text{Slope} = -\frac{1}{8}$

$y - 9 = -\frac{1}{8}(x - 1)$

The Normal Distribution



Emperical Rule: 68% / 95% / 99.7%

68% of the data is within one σ of the mean.

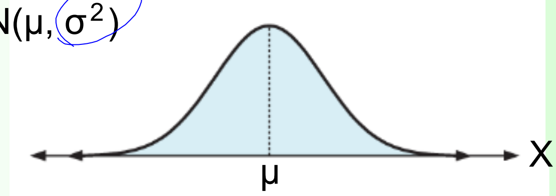
95% " " two σ's "

99.7% " " three σ's "

Using Notation

Variance

$X \sim N(\mu, \sigma^2)$



The random variable X is normally distributed with a mean, μ, and standard deviation, σ.

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

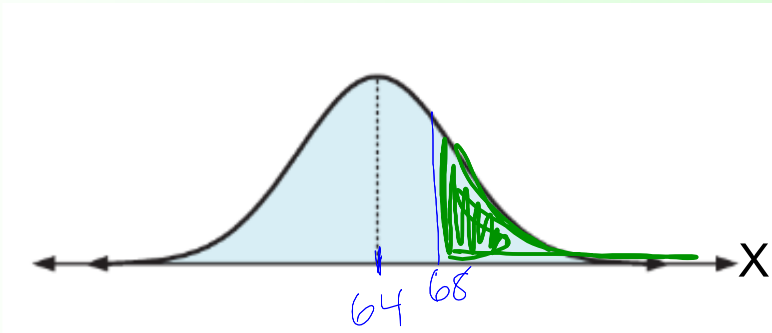
$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

Example

$X \sim N(64, 4^2)$

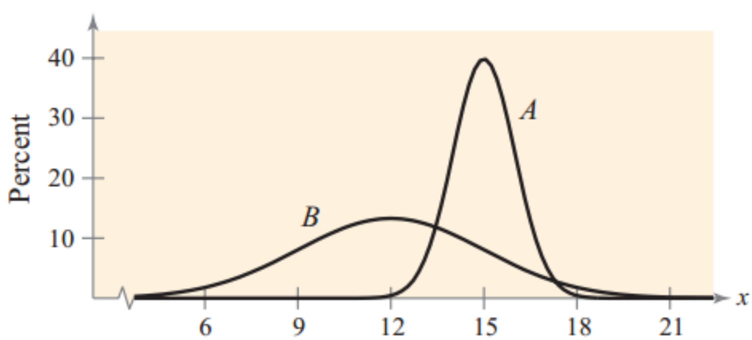
μ σ^2

Show: $P(X > 68)$



Which normal curve has a greater mean? A

Which normal curve has a greater standard deviation? B

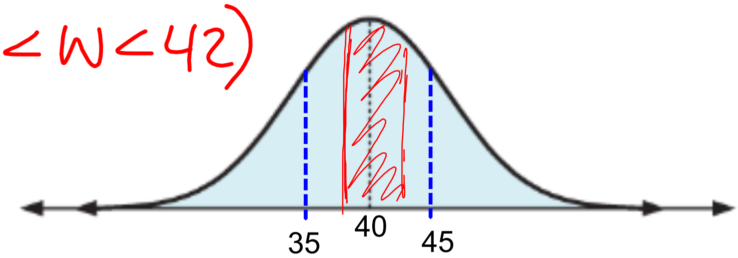


Calculating probabilities with the grapher

Suppose the weights of a bag of potatoes are normally distributed with an average weight of 40 lbs and standard deviation of 5 lbs.

What is the probability that the next bag you buy will be between 38 and 42 lbs?

$P(38 < W < 42)$



2nd VARS DISTR

$P(38 < W < 42)$

normalcdf (38, 42, 40, 5)

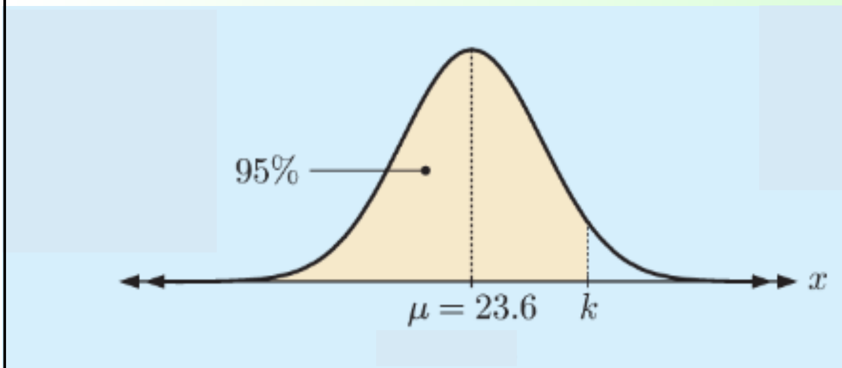
lower boundary

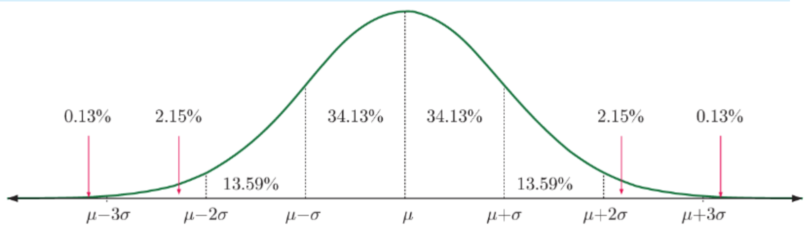
upper boundary

μ σ

≈ 0.311

$X \sim N(23.6, 3.1^2)$ $P(X \leq k) = 0.95$





Calculating Inverse Normal

to find k, given: $X \sim N(23.6, 3.1^2)$

$P(X \leq k) = 0.95$

2nd DISTR VARS

invNorm (0.95, 23.6, 3.1)

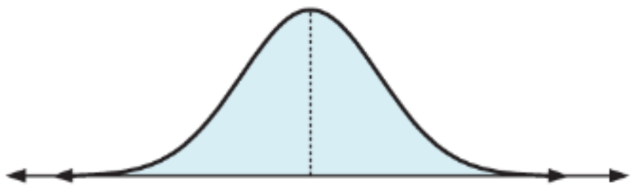
μ σ

probability

$k \approx 28.7$

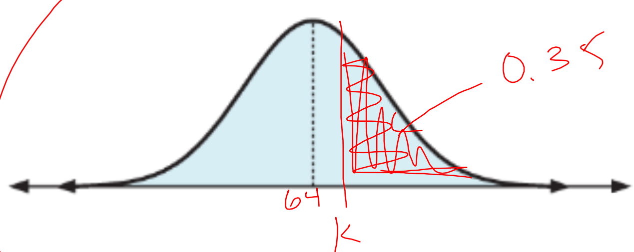
Another Example, illustrate and find k

$X \sim N(64, 3^2)$ $P(X > k) = 0.35$



Another Example, illustrate and find k

$X \sim N(64, 3^2)$ $P(X > k) = 0.35$



$= P(X \leq k) = 0.65$

InvNorm(0.65, 64, 3)

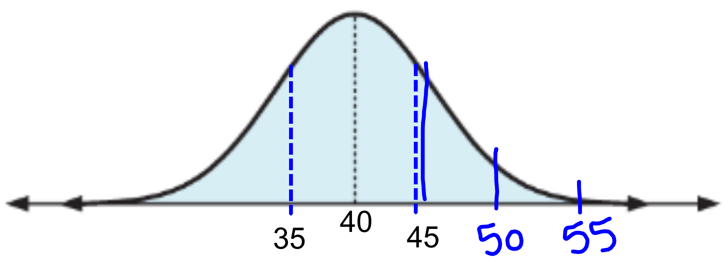
$k \approx 65.2$

How about P(X > 46) ?

normalcdf (46, 100, 40, 5)

σ

What is a reasonable upper boundary?



How about P(X > 46) ?

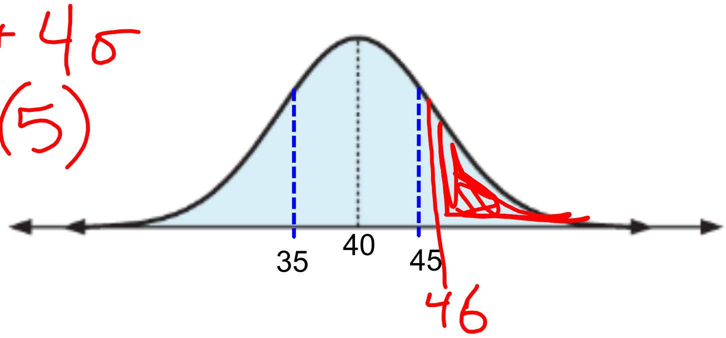
normalcdf (46, 60, 40, 5) $\approx$ 0.115

What is a reasonable upper boundary?

$40 + 4\sigma$

$40 + 4(5)$

60



Two Variable Statistics

Relationship between variables

Using pairs of Numerical data

Scatter Plots
Correlation
Line of Best Fit
Modeling

Using two Categories of data

Chi-Square Test of Independence

X^2

The Chi-Square Test of Independence

Vocabulary...

Independent: When 2 variables are independent, one does not affect the other. The probability of one has no influence over the probability of the other.

H_0 = Null Hypothesis

A statement that the data sets are independent

H_1 = Alternate Hypothesis

A statement that the data sets are not independent

Matrix: An organized array of numbers

Contingency Table: A matrix arrangement of catagorical data.

Entries in the table represent the observed *__number of__* *occurances (not percents). The entries are frequencies, so they are natural numbers,* N, *(counting numbers)*

The probability of 2 Independent Events both happening = the product of the 2 probabilities.

	Graduated	Failed to Graduate	Total
Experimental	73	12	85
Control	43	39	82
Total	116	51	167

in total population

The probability of being in the Experimental group is:

$\frac{85}{167}$

The probability of someone graduating is: $\frac{116}{167}$

If the events are Independent, then the probability of being BOTH in the experimental group and a graduate is:

$\frac{85}{167} \cdot \frac{116}{167} \approx 0.354$

	Graduated	Failed to Graduate	Total
Experimental	73	12	85
Control	43	39	82
Total	116	51	167

Now that we know the probability of being in both groups is: 0.354

How many students would we expect to be a graduate from the experimental group?

$$\left(\text{Prob. of both}\right)\cdot\left(\text{Total population}\right) = 0.354(167) \approx \boxed{59.1}$$

= Expected number of graduates in the experimental group if the events are Independent.

Expected Table of Values

The calculated values you would expect if the variables are Independent

Contingency Table:

	Graduated	Failed to Graduate	Total
Experimental	73	12	85
Control	43	39	82
Total	116	51	167

Expected Value Table:

	Graduated	Failed to Graduate	Total
Experimental	A = 59.1	25.9	85
Control	56.9	25.1	82
Total	116	51	167

$$A = \underbrace{\left(\text{Prob. Exp}\right)\left(\text{Prob Grad.}\right)}_{\text{Prob of both.}}\left(\text{total pop.}\right) = \text{Expected \# of students who grad from Exp}$$

You try:

EXERCISE 11E.1

1 Construct an expected frequency table for the following contingency tables:

a

	Likes chicken	Dislikes chicken	sum
Likes fish			60
Dislikes fish			40
sum	75	25	100

P likes fish → $\frac{60}{100}$

& likes chicken $\frac{75}{100}$

EXERCISE 11E.1

1 Construct an expected frequency table for the following contingency tables:

a

	Likes chicken	Dislikes chicken	sum
Likes fish	A = 45	15	60
Dislikes fish	30	10	40
sum	75	25	100

total pop.

$$A = \underbrace{(\text{Prob. likes fish})(\text{Prob. likes Ch.})}_{\text{Prob. of liking both}}(\text{Total pop})$$

$$A = \frac{60}{100} \cdot \frac{75}{\cancel{100}} \cdot \cancel{100}$$

$$= 45$$

You won't need to do this!

Calculating X^2 by hand:

$$\chi^2_{calc} = \sum \frac{(f_o - f_e)^2}{f_e}$$ where f_o is an observed frequency and f_e is an expected frequency.

f_o original data from contingency table

f_e from your calculated expected values if the variables are assumed Independent

With calculator:

Enter f_o values into a matrix.

2nd MATRIX ► EDIT, ENTER

Enter # of Rows by # of Columns, then your table values without the sums.

Run the X_2 test: STAT ► TESTS ▼ X^2 test

The calculator will automatically calculate the expected frequencies, store them in another matrix and run it all through the formula!

DEGREES OF FREEDOM: The number of values that are free to vary

Contingency table:

	A_1	A_2	*sum*
B_1	9	3	12
B_2	6	2	8
sum	15	5	20

The value in the top left corner is free to vary, as it can take many possible values, one of which is 9. However, once we set this value, the remaining values are *not* free to vary, as they are determined by the row and column sums.

df = (# of rows - 1)(# columns - 1)

df = (2 - 1)(2 - 1)

df = 1 (for any 2X2 table)

For a 3X3 table:

	C_1	C_2	C_3	*sum*
D_1	5	3	4	12
D_2	2	4	2	8
D_3	6	2	5	13
sum	13	9	11	33

df = (3 - 1)(3 - 1)

df = 4 values that are free to vary, the rest will be found by subtraction so they will not vary.

SIGNIFICANCE LEVEL

As the χ^2 value gets larger, it becomes increasingly unlikely that the variables involved are independent.

The **significance level** indicates the minimum acceptable probability that the variables are independent.

The commonly used Significance Levels are 1%, 5%, and 10%

When X^2 > the value determined by the significance level chosen, we reject our Null Hypothesis.

Critical Value of X^2

The value determined by the significance level chosen

When X^2 > Critical Value,

we reject our Null Hypothesis, H_o.

Degrees of freedom (df)	*Significance level* 10%	5%	1%
1	2.71	3.84	6.63
2	4.61	5.99	9.21
3	6.25	7.81	11.34
4	7.78	9.49	13.28
5	9.24	11.07	15.09
6	10.64	12.59	16.81
7	12.02	14.07	18.48
8	13.36	15.51	20.09
9	14.68	16.92	21.67
10	15.99	18.31	23.21

Example:

5% Significance Level

df = 4 (3X3 table)

Critical Value of X^2

= 9.49

Degrees of freedom (df)	*Significance level* 10%	5%	1%
1	2.71	3.84	6.63
2	4.61	5.99	9.21
3	6.25	7.81	11.34
4	7.78	9.49	13.28
5	9.24	11.07	15.09
6	10.64	12.59	16.81
7	12.02	14.07	18.48
8	13.36	15.51	20.09
9	14.68	16.92	21.67
10	15.99	18.31	23.21

H_o rejected

if our calculated X^2 > critical value from table

>9.49

Review:

Chi Square Test of Independence.

χ^2 is a statistic that measures the difference between observed values and expected values in a contingency table.

If the calculated Chi Square value is big enough, we can establish a:

link between two variables

association between two variables

relationship between two variables

Independent ⟷ Not Independent

~~Dependent~~

The test is for Categorical variables only.

- Set up a contingency table for two categorical variables.
- Assume independence to start
- Calculate expected values

We only use the term "correlation" with with numerical data.

The cutoff, or critical Chi-Square Value, is either given to you or found in a resource table.

This value will tell you whether to accept or reject the assumed independence between the two variables.

THE FORMAL TEST FOR INDEPENDENCE

Step 1: State H_0 called the **null hypothesis**. This is a statement that the two variables being considered are independent.
State H_1 called the **alternative hypothesis**. This is a statement that the two variables being considered are not independent.

Step 2: State the **rejection inequality** $\chi^2_{calc} > k$ where k is the **critical value** of χ^2.

Step 3: Construct the expected frequency table.

Step 4: Use technology to find χ^2_{calc}.

Step 5: We either reject H_0 or do not reject H_0, depending on the result of the rejection inequality.

Step 6: We could also use a ***p*-value** to help us with our decision making.
For example, at a 5% significance level: If $p < 0.05$, we reject H_0.
If $p > 0.05$, we do not reject H_0.

If the p-value is smaller than the significance level, then it is sufficiently unlikely that we would have obtained the observed results if the variables had been independent. We therefore conclude that the variables are not independent.

p-value: probability of getting observed values as far, or further, from the expected values (assuming independence)